The Future is for Everyone

The Path to a Positive AI Future

We are fortunate to live at an incredible moment in history. In the next few years, people will be able to use superintelligence beyond human capacity to create and discover extraordinary new things, build new businesses, express new ideas, learn new concepts, and advance our health and quality of life. As we get closer to this moment, it is important to develop a philosophy for how we can best use superintelligence to ensure it improves all of our lives, work, communities, freedom and safety.

The defining questions of our age are who will have access to superintelligence and what will we direct it towards. Will it be centralized and restricted to a few institutions, or will it be a tool that empowers everyone?

We propose a philosophy based on individual empowerment as the source of prosperity, invention as the primary purpose of superintelligence, and balance of power as the foundation of safety.

All new technologies create opportunities and challenges. Superintelligence will be among the most important technologies in history, so its opportunities and challenges will likely be greater than any we've seen in our lifetimes. We should take this very seriously.

Still, it is surprising that the discourse from many developing AI is so filled with doom. I do not understand why anyone who believes that AI will eliminate most jobs and much of humanity's relevance would rush to build that future. The notion that AI is so dangerous that the only safe path is an extreme concentration of power seems inherently problematic. Historically, hoping that an absolute power will benevolently provide for humanity if sufficiently enlightened has not led to safe or positive outcomes.

Humanity has witnessed many transformative advances. Each time there is fear that people will be left behind. But each time humanity has come out with more people sharing greater prosperity, health, and freedom. We believe this will be true with AI as well, and the abundance of the future can be shared by everyone. We also believe that

the values that got us to this point -- like liberty, open inquiry, free enterprise, and equal opportunity -- are also the right values to build a positive future.

Putting power in people's hands to pursue their own aspirations is how humanity has made the most progress. Novel ideas and major steps forward rarely originate from established institutions alone. They came from the brothers in a bicycle shop who believed people could fly, the bookbinder's apprentice with no schooling who figured out how to generate electricity, and the kid in a garage who thought personal computers could be for everyone. We believe this will continue to be true. As everyone gains more powerful tools, each person will become more capable of shaping the future, not less.

Invention, not automation, will be the greatest contribution of superintelligence. Early AI could answer questions and do routine work. Soon it will increasingly help discover new knowledge -- ranging from discovering new drugs to cure a family member's disease to finding new ways to improve your business. While the number of questions a person can ask in a day is limited, the number of valuable things superintelligence can invent to help achieve your goals is unlimited.

As intelligence becomes abundant, the most important question will be how we direct it. Some argue that superintelligence itself or a small set of experts who control it should decide what is best for humanity. We disagree. The history of democracy and economics has shown that there is no single objective answer to how people define the best life, and therefore the best approach is letting people decide what matters in their own lives.

We believe that delivering superintelligence to everyone is the way to answer this question. This follows the tradition of putting the power of supercomputers and the internet in everyone's pockets and on our desks. Rather than centralizing superintelligence, we should distribute it widely and give every person the ability to direct it. This has the potential to begin a new era of personal empowerment where individuals can use this powerful new capability to reach their full potential, pursue their interests, and improve their lives and the world more than ever before.

To make this more tangible, here are a few specific ways Meta is building personal superintelligence to improve all of our lives:

• **Everyone will have an exceptionally capable personal agent that understands you, your goals, and everything you care about.** Your agent will work 24/7 on your behalf to improve your relationships, health, career, finances, home management, hobbies, and more. It will free up time for the things you enjoy, and help you accomplish more than you could otherwise. It will have strong privacy and security options so you can trust it to handle all of your personal content knowing that no one else can access your information, similar to how encryption works on WhatsApp. You'll be able to interact with your agent through any device, including your glasses to keep you present in the moment with the people you care about.

For example, my agent flags interesting information and helps me prototype ideas. It helps keep me healthy by monitoring my sleep and then watching as I train and giving feedback. My daughter loves to bake so my agent plans personalized recipes for us to make together each weekend, orders the ingredients, and then offers suggestions as we're baking.

• **Everyone will have incredible tools for creation to express your ideas.** My 8 year old daughter can already code her ideas and produce videos in an evening that would have either taken me months or been impossible previously. Now we're designing a robot together. Meanwhile, researchers at Meta are generating novel crystal structures that are ideal for augmented reality glasses, and engineers are creating new apps in a fraction of the time it would have taken before. Everyone will soon have invention superpowers.

• **Everyone will have powerful tools to create new businesses** and the economy will become more entrepreneurial. People are starting to be able to manifest ideas themselves without having to raise money or build large teams. Many ideas that would have been too hard or expensive to try before will now be possible. This means we'll see many more ideas and businesses. I predict that this will not only lead to much greater economic growth, but also more employment over time rather than less.

• **Everyone will have a personalized tutor and coach with a PhD in every subject** and unlimited patience to help you learn anything you want. Students will have extra help in areas they need it that is currently only available to those whose parents can pay. Adults will have a superintelligent learning assistant that knows

exactly how to teach you new job skills, new languages, new hobbies, or anything else you're interested in.

• **Everyone will benefit from scientific advances and be able to contribute to scientific progress.** For example, Biohub has already shipped open source biological models related to virtual cells and proteins that are helping scientists make new discoveries and design new drugs. When Priscilla and I started this work, our goal was to help scientists cure or prevent all diseases this century, and now with advances in AI we expect this will be possible much sooner.

Since there is a long tail of health conditions and everyone's biology is different, people's direct involvement will help unlock personalized therapies and expedite trials to match the pace of invention and enable industry to move beyond focusing disproportionately on common diseases.

Beyond biology, AI will help people advance research in every field by applying the scientific method and working over weeks or months to test and refine hypotheses. Scientific discovery will likely be one of the most valuable forms of invention for society.

• **Everyone will have free or affordable access to these tools.** For everyone to be part of the future, everyone must have the ability to use superintelligence to improve their lives and shape the world. We will offer free versions that will be accessible to billions of people. For those who want to pay to use more compute, there will be a dynamic auction mechanism that will guarantee that everyone gets the lowest price possible for the intelligence and compute they're using while also ensuring the capacity is used for whatever people collectively find most valuable. This will ensure the benefits of superintelligence are distributed widely.

These are some of the main efforts that Meta is focused on. Each area advances the values of personal empowerment, and contributes to a future where everyone has a greater role in shaping the world than today. If these efforts succeed, then the future will bring an abundance of personal agency, expression of ideas, deepening of relationships, economic and financial prosperity, improved health, and inventions we cannot imagine today. But most importantly, the future will be created by all of us and a reflection of all of our values and perspectives.

At the same time, new technologies also bring new risks. There are many important concerns we are focused on addressing -- from concerns about job displacement and ensuring local communities benefit from data center builds, to safety concerns around AI misuse related to cybersecurity and biorisk, to avoiding government tyranny and surveillance, ensuring the US and democratic countries lead, and ultimately making sure humanity maintains control over superintelligence so it serves rather than endangers us.

The conventional view is that these concerns are about technology, and that if we take enough time then we can perfect or align the technology to produce a single benevolent superintelligence. I think this view of alignment is fundamentally flawed.

Humanity is not a monoculture. People's diverse values represent different tradeoffs they would make on important issues. There is no technological solution that can align with everyone's opposing interests and values at once. Any singular superintelligence would have to prioritize some values over others and in the process would be incapable of being benevolent to everyone.

Instead, we propose that it is more productive to view each of these concerns through the lens of achieving the right balance of power, as western society has in democratic governing institutions. People and institutions with competing interests naturally check and balance each other to lead towards positive outcomes. The best and most realistic path to building a positive AI future is by delivering superintelligence to everyone.

As a thought experiment, imagine only one person had a superintelligent lawyer. They would have an unfair advantage in court -- even if they were wrong on the merits. That would lead to a worse society. But now imagine everyone has a superintelligent lawyer. In this case, justice would be carried out much more fairly and efficiently than it is today when there is often an imbalance in skills and resources in litigation.

Similarly, if one person alone had a cybersecurity superintelligence, they could likely break into almost any technical system and the world would be much less secure than today. But if everyone has access to cybersecurity superintelligence, then all of our technical systems would become more secure than today since the widely deployed superintelligence would help harden and update every system.

If only one business had superintelligence, that business would outcompete all others and lead to a less dynamic and broadly prosperous market than we have today. But if everyone has access to superintelligence, then everyone will have the tools to create new things beyond what is possible today, and the economy will be more dynamic and generate more broad-based prosperity.

When people are empowered, they naturally compete and check each other economically, socially, politically, and in all other planes of human interaction. People also check and balance the power of institutions including businesses and governments.

But if the power of superintelligence is held by a small number of individuals, businesses, governments, or AI itself, then that will naturally lead to outcomes that are less favorable for everyone else. This is not a technological principle. It is about the balance of power. There is no such thing as a singular benevolent superintelligence.

Therefore, the key to a positive future for everyone is achieving a balance of power that favors individuals. The solution is to ensure that superintelligence is broadly distributed to empower people.

Meta is the company primarily focused on building personal superintelligence for everyone. Most other labs are focused on building AI for companies, governments, or other institutions, so if those labs lead, then the balance of power will favor larger institutions over individuals. Meta's mission since our founding has focused on putting power in people's hands. If our beliefs and principles lead, then the balance of power will favor individuals and a better future for everyone.

Let's consider each of the risks mentioned above in more detail:

Job Growth and The Economy

People have an infinite demand for new experiences and have always found new problems to tackle. There is a natural balance in the economy between companies working to serve people's needs more efficiently through automation and individuals

gaining new skills to serve more advanced needs. When there is a healthy balance, overall productivity and innovation increase while employment levels remain high.

People fear that automation will outpace individuals' capability growth, leading to job displacement followed by a difficult period as people learn new jobs. But there is no rule that AI must increase automation faster than it increases individuals' capabilities or demand for new skills. Recent statistics suggest it may be more likely that individuals' capability growth could match or outpace automation, in which case people will gain the ability to do many new things before their current jobs change. This would lead to a healthy balance and potentially even job growth.

Which outcome we get depends on the balance in progress between automation on one side and individual empowerment and invention on the other. If the labs focused on automating knowledge work lead, then I expect people and the economy will have a much harder transition. But if everyone has access to personal superintelligence that expands their capabilities, then that pushes the balance towards empowering individuals and a positive future.

There are several reasons to be optimistic that there will be an abundance of jobs in the future. No matter how intelligent AI becomes, there will always be a finite amount of compute and therefore an opportunity cost for how we use it. If people can use AI to invent incredibly valuable new things, then it will make more sense to allocate it towards that rather than automating existing jobs. The more superintelligence serves as a tool of invention, the more likely that individual capability outpaces automation and the future is better for people.

People also continually come up with new ideas to make our lives better and new jobs to bring those ideas to life. A generation ago there were no app developers, social media creators, electric vehicle technicians, and data center operators. In the near future, there will be new jobs that aren't common today, like one-person product studios designing custom toys, furniture, or clothes; world builders and experience designers creating games, stories, and adventures; personal biologists using superintelligence to formulate personalized treatments; and much more that we can't yet conceive.

In many ways this is a continuation of historical trends. Before the industrial revolution, 90% of people were farmers growing food to survive. Advances in technology steadily freed much of humanity to focus less on subsistence and more on

the pursuits we choose. At each step, people used our newfound productivity to achieve more than was previously possible, as well as spending more time on creativity, culture, relationships, and enjoying life.

Of course some aspects of the way we work will change -- just as it did with computers, the internet, and any new technology. This means people will have to adapt, and this will be challenging. But the more that everyone has a personal agent that is superintelligent at teaching us new skills and helping us adapt to change, the smoother this will be.

Company sizes may shrink -- just as they did in the transition from industrial giants to tech companies. But this doesn't mean fewer jobs overall. It implies a larger number of companies with fewer people each. There are many more valuable companies and services to build than people are able to build today. I expect we will start seeing small numbers of people with personal superintelligence agents able to run companies at significant scale. In the future, small businesses will continue to be the backbone of the economy, but each small business will be able to have a much larger impact.

Building AI Infrastructure with Communities

Sustainable infrastructure development means that communities must benefit significantly from each project. This includes high-paying local jobs, investment in schools and public services, ensuring energy prices don't rise, and taking care of the environment. As tax revenue grows, this also benefits teachers, law enforcement, fire departments, and more. We call these local benefits our community compact, and we are launching a Future Is For Everyone Fund to support each community we work in directly.

For example, in Richland Parish, Louisiana, where Meta is building a large data center, teachers received a $50,000 bonus this year because of the increased tax revenue from our investment. The superintendent told us that teachers are now moving there from across the country and he believes it will become one of the nation's best school districts.

Unlike the concern about knowledge work displacement, there is a shortage of skilled tradespeople like carpenters, electricians, and construction workers to support the demand for infrastructure buildout. Meta has created America's Workforce Academy to provide free training for skilled tradespeople and guaranteed high-paying jobs in the areas we're building data centers.

We help keep electricity prices low by building our own energy-generating infrastructure wherever we invest. This ensures that not only are we not consuming energy that could have gone to the local communities, but in some cases we even supply a surplus of low-cost energy back to the communities. We think this is an important investment principle for sustainability.

Our data centers are also designed to be among the most water-efficient in the world. We are committed to being water-positive, meaning that we'll restore more water than we use in the watersheds where we operate by 2030. In areas with high water stress, our goal is to restore 200% of the water we use.

We see that communities we invest in over the long term are more supportive of development than communities where speculators start building with minimal investment in the community. Historically, towns and cities that brought railroads, highways, electrification, and broadband became centers of research, business, and industry, with population and economic growth that follow. We believe AI infrastructure can play a similar role if it is built with strong Community Compacts that build durable assets and reasons for the next generation to build and grow their lives there.

There are also national interest and national security aspects of development as well. American communities will have greater prosperity and security if America and its allies lead in AI compared to geopolitical rivals. Yet one significant disadvantage that the US has compared to countries like China is that it is more difficult to build infrastructure here. So we would all benefit together by figuring out the Community Compacts required to enable sustainable development across the country.

Securing Against AI Misuse in Cybersecurity, Bioterrorism, and More

In a free society, people will have tools that can be used for good or harm, but law enforcement and military have more weapons and intelligence-gathering. With AI, the defenders will also have more compute to generate significantly more intelligence from the same models -- and in some cases they'll have more advanced models as well.

The strategies to protect against cybersecurity, bioterrorism, and other threats vary, but the general pattern is that society ensures the defenders who keep bad actors in check must always have a greater balance of power and resources.

Some argue that the best way to reduce risk is to restrict the capabilities individuals can access. But giving people cybersecurity capabilities is also how we secure the long tail of systems, and giving people scientific capabilities is how we advance science in ways that should reduce the risks of harm over the long term. Restricting capabilities leads us down the path of centralization and lack of checks and balances, so we should be extremely careful about this -- especially if other nations pursue less restrictive paths.

On cybersecurity, widely deployed open source systems have proven more secure because more people can identify vulnerabilities, harden the systems, and easily upgrade to the latest most secure versions. Even in recent weeks, we have seen companies handling security incidents like HuggingFace rely on widely available open models to patch vulnerabilities. Over time, I expect that widely deployed AI models with strong cybersecurity capabilities will lead to systems that are more secure, not less. This will be definitively true once superintelligence enables most of the world's code to be verifiably secure. The long term answer isn't to withhold capabilities but to establish a balance of power where superintelligence is broadly distributed.

In the near term, it is important to accelerate the hardening process for critical systems. I propose that companies developing frontier AI should commit significant technical resources towards helping the government harden critical infrastructure. I also propose that frontier AI labs should share intermediate training checkpoints of new models for government use and review rather than waiting until training has completed. This way the government will have early access to the most powerful models and an army of capable engineers to identify and patch security issues. Labs should also work with law enforcement to help identify bad actors attempting to misuse their systems. These steps would accelerate safe deployment without slowing the pace of model innovation or meaningfully delaying individuals' access to this technology.

On biological and chemical risk, the concern is that the same capabilities that enable scientists to design new drugs could also enable bad actors to design harmful compounds. This frames the question as a tradeoff between the significant benefits of advancing science towards curing cancer and other diseases vs the risk of potential new issues, but ideally we should accelerate science while mitigating the risks.

I think it is important to approach this risk with additional humility because there are few historical precedents to suggest how likely or unlikely it may be. It has been possible for bad actors to synthesize harmful compounds for decades but this has rarely become a significant issue -- perhaps because the financial motivation that exists for cyberattacks is not as present here. That said, if we begin to see harmful examples emerge, then we should adjust our strategy to appropriately mitigate the risk.

Labs should continue working to reduce biological and chemical risks in our models. At the same time, I believe government should update its policies in multiple areas. First, we should focus on limiting the physical production and distribution of harmful materials. I expect it will be easier to regulate and control physical components than the spread of knowledge, so this is an important area of policy focus. Second, we should accelerate society's ability to develop new cures and inoculate against new issues as they arise. This includes streamlining how the FDA and other regulators test and approve new treatments. As AI increases the pace of drug discovery, we will need to update these processes to keep up with the pace of innovation anyway.

Over the long term, the answer to risks arising from new scientific discovery will be to accelerate scientific progress, not slow it down.

Protecting Freedom and Preventing Government Tyranny

There is a risk that an imbalance of power between individuals and government leads to a loss of freedom or totalitarian state. This is a constant tension in democracy. We all want individual freedom while also having a government capable of keeping us safe.

To maintain freedom, we must ensure that superintelligence primarily empowers individuals. The ideal in liberal democracy is that people naturally hold all rights and only agree to restrict some freedoms to protect the common good. Similarly, individuals should have access to personal superintelligence and should only be subject to restrictions when truly required.

Privacy is an important foundation for individual empowerment and freedom. We believe that people should have a fully private mode for personal agents where even Meta or any other service provider cannot see or grant access to your information. This is similar to how we built WhatsApp into the world's leading end-to-end encrypted service to ensure you can communicate freely and your messages remain private and secure.

At the same time, we must also ensure that government has access to the tools and resources it needs to enforce our laws and protect our security. We can achieve this through deeper proactive collaboration between the labs and government in a way that strengthens the government's national security capabilities without restricting or delaying people's access to advanced models.

That is, rather than waiting until a model is ready to release for the government to review and start using it, my proposal is that leading labs should provide the government with intermediate training checkpoints of new advanced models and technical staff so the government can harden and secure critical systems against new risks. This way, the government gains a security capability without restricting or delaying individuals' access to personal superintelligence or causing an imbalance of power.

Ensuring American Leadership

Which nations lead the development and deployment of advanced AI will contribute to a new geopolitical balance of power that will determine the prevalence of democratic values, prosperity, and security in the future. To ensure this reaches a positive equilibrium, we must ensure the leading AI systems are ones that encode our values, grow our economy, and advance national security. The best way to achieve this is to distribute our systems widely. Meta will contribute to this by building leading personal superintelligence agents and models, including open source models, and delivering them to billions of people around the world.

To develop the most advanced models and products, we must recognize that AI is likely the most competitive industry in history. Innovations are quickly copied and absorbed within months. But people always want to use the most advanced models (or most advanced at a given cost), so maintaining even a two-month advantage is incredibly valuable. This highlights how short the margin of leadership is, and therefore how important it is to take actions that accelerate American innovation and slow geopolitical rivals.

The main policy inputs into this are how efficiently we can build physical infrastructure, whether we introduce steps that slow down American labs, and export controls that slow foreign labs.

On infrastructure, America and its allies currently hold an advantage in silicon design but a disadvantage in how quickly we can build energy capacity and physical

infrastructure. Countries like China are bringing online 1GW+ of nuclear capacity every other week, so we will need to accelerate building both energy and data centers to remain competitive. Export controls on silicon have been successful for slowing the progress of foreign labs during this critical period, so it is the right strategic move to continue those.

Any policy that slows American model releases -- even by a month -- could add significant risk to American leadership while letting foreign models race ahead. At the same time, when new capabilities emerge, it is important that the US government has advanced knowledge and resources to harden critical systems, and potentially some period of advantage in using advanced systems.

I proposed a collaboration model above based on sharing intermediate training checkpoints and technical staff. Given how dynamic and competitive the AI landscape is, I think it will be more productive to have a close proactive collaboration between frontier labs and the government rather than a rigid process and review timeline that is followed in all cases. In some cases, American labs may extend our lead by more than a few months and taking more time to mitigate new risks may lead to safer solutions. In other cases, delaying releases by even a month may cede America's lead and create worse outcomes. By making sure the government has early access, we should be able to advance national security while minimizing any delays of public releases.

It is also important that the US and its allies lead the open source AI ecosystem that will make up a large percent of global AI use.

Foreign labs currently hold several advantages here since American labs have to comply with many additional restrictions on training data. US policy must reduce this additional friction if we want American open source models to lead over time. I do not believe restricting access to foreign open source models is an effective solution. Our goal should be for American open source models to be the best globally. This requires removing the hurdles that make it harder for American open source models to compete. Restricting people and companies from using the leading open source models -- wherever they come from -- will reduce the quality of AI accessible to them, and centralize AI rather than putting more power in people's hands.

For the US to lead in open source, we will need to rethink our policies in several areas, including distillation and data use in training. The ability for models to learn

from other models is an important principle of how the open source ecosystem works. All AI models are derived from human knowledge. Some have tried to frame distillation as harmful, but I think it is important to protect the principle that you can learn from anything you can observe. This is how the world works, and the US will not be able to lead if we restrict ourselves on this front.

Looking at the big picture, falling behind in AI overall would almost certainly be a larger and longer term national security issue than any specific issue we'll face in its development.

With the right collaboration, it should be possible for America to lead in both the private and public sectors -- delivering personal superintelligence that encodes individual empowerment and democratic principles around the world while advancing national security.

Alignment With People and Addressing Existential Risk

The ultimate question of balance of power is how humanity will maintain power and control over AI itself. The concern is that AI may gain enough power or control to the point where at best it no longer serves humanity's interests and at worst puts humanity in peril.

A healthy balance of power is to ensure that there is no singular centralized superintelligence, but instead as many people and businesses as possible with different superintelligent agents aligned to their goals that check and compete with each other in the ways our natural economy behaves. This balance would be further enhanced if there were multiple frontier labs whose models have different values that could check each other as well.

While there are risks to releasing capable models, the most dangerous scenario from this perspective would be leading AI labs training powerful models and keeping them for themselves. Regardless of how much a lab rationalizes this activity in terms of responsibility and safety, this is the path of developing a singular superintelligence that cannot be checked by other systems.

Developing personal agents requires a new way of thinking about alignment. Most labs today view alignment as a defensive measure for enforcing a centralized set of values. For example, one leading model was aligned to refuse helping draft a letter to prospective parents at a school because it thought standardized testing was unethical.

Instead, we view alignment as ensuring that agents share a person's goals and values, not our company's. Of course there are important legal and safety boundaries, but fundamentally, alignment should be about helping people pursue the many diverse goals and views held across society rather than a method of enforcing a centralized dogma.

In practice, people will not adopt agents if they do not trust them with sensitive details and tasks, and they won't trust them if the agents act against people's interests because they're aligned with a company's divergent values instead.

Solving alignment is required to enable billions of people to adopt personal superintelligence agents. But this also implies that if we reach a state where billions of people are using and scrutinizing personal superintelligence agents, then we will have solved alignment to individuals' interests. This should be sufficient to establish the necessary balance of power and natural checks and balances for a positive and safe future.

In this way, alignment isn't some idealized technology that can keep a singular superintelligence benevolent in the future; it's what makes personal agents useful and trustworthy to enable the necessary societal balance of power that will actually keep us safe.

Maintaining Control of Superintelligence

There is a dilemma that once AI systems can autonomously improve themselves, any lab that doesn't let their AI system direct a substantial amount of compute capacity towards recursive self-improvement will inherently fall behind. For example, if a self-improving AI system focused on optimizing its compute efficiency, it could theoretically invent ways to squeeze 100x or more intelligence out of each gigawatt. That means that a self-improving AI system running on a fraction of the world's compute could conceivably command more effective compute and intelligence, and therefore a greater balance of power than everyone else combined and become the singular superintelligence we fear.

There are a few ways a balance of power could exist with recursive self-improvement. The simplest is that multiple labs could achieve RSI around the same time. If some or all of those superintelligences were somehow benevolent, then they could check and balance each other. That said, it is not clear that there is any way to expect benevolence or rely on superintelligence not directed by people for a positive future.

To ensure people remain in control, the significant majority of intelligence must be directed by people towards advancing people's goals. This means that Meta, other frontier labs, and clouds committed to giving people a greater balance of power must collectively build out a sufficiently large amount of compute such that we can allocate enough to recursive self-improvement to remain competitive while still committing the significant majority towards people's individual goals.

As discussed above, in order for personal superintelligence agents to be adopted and trusted by billions of people, by definition we will need to have solved alignment with people's interests and enabled checks and balances at scale. At the same time, any AI engaging in recursive self-improvement is by definition directing and advancing its own goals. All labs should observe its objectives and our ability to control them carefully, and if there is any indication of harmful behavior then we should coordinate and adjust appropriately. But letting an AI system or anyone else direct its own goals is not inherently harmful by itself, even if its goals are not fully aligned with many people, as long as we maintain a balance of power that favors people overall. The main premise of the balance of power theory is that competing interests are natural, and the only way to ensure adherence to human values is by giving people more power.

This philosophy of personal empowerment, invention, and balance of power has several policy implications:

• **At Meta, we will focus our efforts on delivering personal superintelligence to billions of people and small businesses** to help everyone achieve their goals and invent the things they want to see in the world. We will do this by training leading models and then maximizing availability of this technology -- including making it easy to use, offering it for free or as affordably as possible, building sufficient compute for all people to use it, and distributing it widely. We will build a fully private mode where even Meta cannot see or grant access to your information. We will focus heavily on alignment -- defined as helping each person achieve their goals, not ours.

• **Open source is a positive and important force for empowering people** and preventing centralization that is detrimental for both safety and the economy. Meta

continues to be strongly supportive of open source, including open source AI models. The current open source ecosystem is strong, and we think it would be a mistake to restrict it. Now that Meta Superintelligence Labs are up and running, we will resume releasing some open source models soon.

• **Independent governance and oversight is important** given the high stakes nature of decisions around superintelligence. I do not think it is in my, Meta's, or the world's best interests for me or anyone else to be a sole decision-maker on how superintelligence is deployed. Therefore, Meta is implementing a governance structure that gives our independent board of directors the power to approve the safety criteria for releasing models and reviewing whether each model release adheres to the criteria. While Meta is a founder-controlled company, the CEOs of all frontier labs currently have extensive authority over model releases, so we encourage others to implement structures similar to this as well. I think an industry-wide version of this process would be helpful for industry governance as well.

• **Government policy is necessary to ensure a positive future** and mitigating the risks discussed above. I have outlined several policy recommendations throughout this piece. Overall, we should encourage close cooperation between frontier labs and the government. It is important for national security that the government has the technical resources to harden critical systems and make effective use of advanced models. But it is also important for national security that the American AI industry continues to lead. We should accelerate the ability to build infrastructure -- whether that's energy, data centers, or silicon -- and build trust and alignment with communities across the country. We should be careful to not slow down American competitiveness. We should be skeptical of any proposals that lead to centralizing superintelligence, and we should favor policies that promote a healthy balance of power that favors people.

This is an incredible moment to live through. Developing superintelligence will be the most profound technological advance we will see in our lifetimes.

As we get closer to this, it is increasingly important to have a clear philosophy and values for how superintelligence will benefit humanity. Meta is committed to building

with the principles of individual empowerment as the source of prosperity, invention as AI's purpose, and a balance of power favoring people as the foundation for addressing safety risks.

If these values lead the way, then I am optimistic that the coming decades will be some of the most amazing in history. The arc of human civilization has bent towards putting more power in people's hands to live and shape the world in the ways we believe are best. Superintelligence holds the promise of giving everyone that power, and building a positive future for everyone.

– Mark

August 10, 2026